TITLE OF THE INVENTION

**System And Method For Reduced Frame Flooding**

CROSS REFERENCE TO RELATED APPLICATIONS

--None--

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

--Not Applicable--

BACKGROUND OF THE INVENTION

The present invention is related to the field of communications network devices, and more particularly to network devices performing frame forwarding in communications networks.

Network devices in communications networks commonly carry out two related functions. One is frame forwarding, by which frames received at a given port are forwarded to one or more other ports for transmission toward their destination. Forwarding is typically carried out using "forwarding tables" existing at each port. The other function is address learning, which enables the device to accurately maintain address information in the forwarding tables for use in performing the forwarding function.

The forwarding of unicast frames generally works in the following manner. When a frame is received by the device, the forwarding table at the receiving port, or "ingress port", is consulted using the destination address (DA) from the frame. If there is an entry in the forwarding table for this address, the frame is forwarded to one other port of the network device as indicated by the entry. Otherwise, the frame is "flooded" to all other ports of the network device, on the assumption that the destination address is reachable via one of the other ports.





ATTORNEY DOCKET NO. FJFNX-183XX
WEINGARTEN, SCHURGIN,
GAGNEBIN & LEBOVICI LLP
TEL. (617) 542-2290
FAX. (617) 451-0313

Flooding is wasteful, however, because all but one of the transmissions from the other ports are unnecessary.

Addresses are learned at ports through "egress learning". As a frame exits the network device at a given port, the port uses the source address (SA) of the frame and the identity of the ingress port at which the frame was originally received to create or update an entry in its forwarding table. The addresses stored as entries in the forwarding table are compared with the DA of a received frame during the forwarding lookup discussed above. Once an address is "known" at a port, i.e., the port has an entry for the address in its forwarding table, the port need not flood frames containing that address.

Physical ports of a network device may be logically grouped into an "aggregated port" (AP), a single logical connection between the device and external equipment such as a bridge. In general, frames can be transferred between the device and external equipment via any of the physical ports of an aggregated port. Generally, a particular port is selected in some deterministic fashion, such as according to a hash function of a network address in a frame.

When an AP is employed, it is possible that frames being sent from a station attached to the AP that are destined for a station attached to another port of the device are transferred via one physical port of the AP, while frames being sent in the other direction between the two stations are transferred via another physical port of the AP. In this case, egress learning as described above may fail to terminate flooding in a timely manner. This happens because the ports use different forwarding tables, and the forwarding table used for forwarding frames in the one direction is not updated with the address information from frames flowing in the other direction. Thus, frames continue to be flooded in a wasteful manner, despite the existence of information



ATTORNEY DOCKET NO. FJFNX-183XX
WEINGARTEN, SCHURGIN,
GAGNEBIN & LEBOVICI LLP
TEL. (617) 542-2290
FAX. (617) 451-0313

at another port of the device that could be used to terminate the flooding.

## BRIEF SUMMARY OF THE INVENTION

In accordance with the present invention, a network device is disclosed that monitors the flooding of frames and obtains information to update the forwarding table at a port to reduce the incidence of unnecessary frame flooding, resulting in more efficient operation of the device and a network in which the device is used. The functionality finds particular use in connection with aggregated ports, but may be used in other configurations as well.

The disclosed network device maintains an unknown address and a count at a first port, the unknown address being a network address for which there is no information at the first port identifying another port of the network device to which frames containing the unknown address are to be forwarded, and the count identifying the number of times frames containing the unknown address have been flooded from the first port to other ports of the network device. Upon the receipt of unicast frames containing the unknown address at the first port while the count is less than a predetermined threshold, the count is incremented and the received frames are flooded to the other ports of the network device. When the count has reached the predetermined threshold, it is determined whether there is information at a second port of the network device identifying a specific port to which unicast frames containing the unknown address are to be forwarded, and if so then the information is transferred from the second port to the first port, whereupon the unknown address becomes known at the first port. Subsequently, upon the receipt of unicast frames containing the now known address at the first port, the received frames are forwarded to only the specific port identified in the transferred information.



ATTORNEY DOCKET NO. FJFNX-183XX
WEINGARTEN, SCHURGIN,
GAGNEBIN & LEBOVICI LLP
TEL. (617) 542-2290
FAX. (617) 451-0313

The disclosed device maintains unknown addresses and corresponding counts in separate bins, and creates and discards bins as necessary to monitor flooding. Timers may be used in connection with the bins to ensure that they are used for more active traffic streams. Thus, if a timer times out before the count in a given bin reaches the predetermined threshold, the bin is discarded. Additionally, if the number of active bins reaches some predetermined maximum, then the forwarding table at the port can be completely re-synchronized with the forwarding table(s) of one or more of the other ports, on the assumption that there may be several useful entries transferred that will obviate the monitoring of multiple addresses residing in the bins.

Other aspects, features, and advantages of the present invention will be apparent from in the detailed description that follows.

## BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING

The invention will be more fully understood by reference to the following Detailed Description of the invention in conjunction with the Drawing, of which:

Figure 1 is a block diagram of a network including a network device in accordance with the present invention;

Figure 2 is a flow diagram of a frame forwarding process as known in the art;

Figure 3 is flow diagram of an address learning process as known in the art;

Figure 4 is a diagram depicting a number of bins for maintaining unknown addresses and counts in the network device of Figure 1; and

Figure 5 is a flow diagram of a process of monitoring frame flooding and obtaining address information to reduce flooding in the network device of Figure 1.



ATTORNEY DOCKET NO. FJFNX-183XX
WEINGARTEN, SCHURGIN,
GAGNEBIN & LEBOVICI LLP
TEL. (617) 542-2290
FAX. (617) 451-0313

## DETAILED DESCRIPTION OF THE INVENTION

In Figure 1, a network element 10 such as a switch is shown having a number of ports 12, labeled "A" through "D". The ports 12 are interconnected by a data communication path 14 that enables the ports 12 to exchange network data frames and control messages. The network element 10 performs frame forwarding and related functions on behalf of a number of client nodes or "stations" 16. Each station 16 is labeled in Figure 1 with a station name, such as "S1", and a station address, such as "M1". The addresses M1 - M4 are media access control (MAC) addresses, such as Ethernet addresses.

As shown, stations S1 and S2 are connected to the network element 10 via a bridge 18, which has two connections 20 and 22 to ports A and B respectively of the network element 10. From the perspective of both the bridge 18 and the network element 10, ports A and B constitute a single "aggregated port" as indicated by a pairing 24 of the connections 20 and 22. In general, frames can be transferred between the network element 10 and the bridge 18 via either port A or B. Generally, a particular port is selected in some deterministic fashion, such as according to a hash function of a network address in a frame. However, the use of either port in particular is largely invisible from the perspective of forwarding software and hardware in the network element 10 and the bridge 18, which treat the two ports as a single logical entity having greater data carrying capacity than either port alone.

The network element 10 carries out two related functions. One is frame forwarding, by which frames received at a given port 12 are forwarded to one or more other ports 12 for transmission toward their respective destinations. Forwarding is carried out using "forwarding tables" existing at each port. The other function is address learning, which enables the network element 10



ATTORNEY DOCKET NO. FJFNX-183XX
WEINGARTEN, SCHURGIN,
GAGNEBIN & LEBOVICI LLP
TEL. (617) 542-2290
FAX. (617) 451-0313

to maintain address information in the forwarding tables for use in performing the forwarding function.

Figure 2 depicts the forwarding of unicast frames. At step 26, a unicast frame is received at a given port 12, termed an "ingress" port, of the network element 10. At step 28, the destination address (DA) is obtained from the frame, and a forwarding table at the ingress port is accessed using the DA from the frame to ascertain the output port 12 to be employed. At step 30, it is determined whether an entry exists in the forwarding table for this DA. If so, then at step 32 the frame is forwarded to another port 12 of the network element 10 as identified by the entry, for re-transmission toward the destination station. In this case, it is said that the address is "known" at the port. If such an entry is not found, the address is said to be "unknown", and at step 34 the frame is "flooded" to all other ports 12 of the network element 10, on the assumption that the destination station is reachable via one of the other ports. As previously mentioned, such flooding of frames is generally wasteful of network switch resources and communication bandwidth, because all but one of the transmissions from the other ports 12 are generally unnecessary. Thus, it is preferred that most searches of the forwarding table result in finding an entry and the forwarding of frames to a single port as shown at step 32.

Figure 3 shows the address learning process, which is referred to as "egress learning". At step 36, a frame is received at a port 12, referred to as an "egress port", from the ingress port via the communication path 14. At step 38, the egress port obtains the source address (SA) from the frame and performs a search of the egress port's forwarding table. At step 40, it is determined whether an entry is found. If so, then at step 42, the entry is updated. That is, the port identification in the entry is re-written with the identity of the ingress port from which the frame was received. In most cases, this updating is unnecessary,



ATTORNEY DOCKET NO. FJFNX-183XX
WEINGARTEN, SCHURGIN,
GAGNEBIN & LEBOVICI LLP
TEL. (617) 542-2290
FAX. (617) 451-0313

because the port associated with the address from the frame has not changed. However, the network or the network element 10 may have been re-configured since the last time frames from the station identified by the SA have been received. In this case, the updating of the entry results in correctly associating the address with a new port 12 of the network element 10 via which the corresponding station is now reached.

If at step 40 no entry is found in the forwarding table, then at step 44 a new entry is created and added to the table. This entry contains the SA and an identifier of the ingress port from which the egress port received the frame. This new entry is thus available in the forwarding table to be used for subsequent forwarding look-ups as described with reference to Figure 2. In particular, subsequent forwarding look-ups for the same address result in the forwarding of the frames to a single port as at step 32 rather than the flooding of step 34.

Referring back to Figure 1, there is a potential problem with the use of an aggregated port. Frames being sent from one station 16 to another station 16 in the network may be received at one port 12 of the aggregated port, while frames being sent in the other direction are transmitted via another port 12 of the aggregated port. For example, it may be the case that frames sent by station S1 to station S3 are received by the network element 10 at port B, whereas frames sent to station S1 from station S3 are transmitted to the bridge 18 from port A of the network element 10. The following table shows how the typical egress learning process such as that shown in Figure 3 can fail to terminate flooding in such a case:



ATTORNEY DOCKET NO. FJFNX-183XX
WEINGARTEN, SCHURGIN,
GAGNEBIN & LEBOVICI LLP
TEL. (617) 542-2290
FAX. (617) 451-0313

| Event | Port A | Port B | Port C | Port D |
|---|---|---|---|---|
| 1. S1 to S3 (flooded by B) | | | (M1, AP) | (M1, AP) |
| 2. S3 to S1 | (M3, C) | | | |
| 3. S1 to S3 (flooded by B) | | | | |

The above table shows the new entries that are created in the forwarding table of each port 12 as a result of each event. For the first transmission from S1 to S3, it is assumed that the frame arrives at port B of the aggregated port (AP). The destination address of M3 is unknown at port B, and therefore the frame is flooded to ports C and D. Each of these ports learns the address M1 and associates it with the aggregated port (AP) through the egress learning process.

For the second transmission, from S3 to S1, port C finds the address M1 in its forwarding table and sends the frame directly to the AP. It is assumed that the outbound frame is handled by port A of the AP. Port A transmits the frame to the bridge 18 for subsequent transmission to station S1, and also learns the address M3 and associates it with port C via egress learning. However, port B of the AP does not learn the address M3.

For the third transmission, port B still has no entry for address M3, and therefore floods the frame again. This occurs despite the fact that the address is known at port A, which is part of the same AP as port B. In the absence of any external correcting mechanism, this situation can exist for a considerable period, resulting in excessive and unnecessary flooding of frames within the network element 10.

It is assumed in the foregoing that each port 12 maintains its own forwarding table largely independently. This may be a generally desirable characteristic, especially when the ports 12 reside on different line cards or otherwise have relatively



ATTORNEY DOCKET NO. FJFNX-183XX
WEINGARTEN, SCHURGIN,
GAGNEBIN & LEBOVICI LLP
TEL. (617) 542-2290
FAX. (617) 451-0313

restricted ability to communicate with each other. In such multi-port network devices, there may be some type of periodic re-synchronizing of the forwarding tables of all the ports 12, or of all the ports 12 of a given AP, and this process can be relied upon to eventually propagate many new or updated forwarding table entries to ports at which they are needed. However, such a mechanism is resource-intensive and is generally performed relatively infrequently. Reliance upon this mechanism alone may result in relatively inefficient operation of the network element 10 and may not reduce the incidence of flooding to a desired degree.

To address this shortcoming of the normal address learning process in connection with aggregated ports, the network element 10 employs processes for monitoring the flooding of frames and sharing information among the ports 12 when the flooding of frames indicates that such sharing may be useful. These processes are described with reference to Figures 4 and 5.

Figure 4 shows a number of bins 46 maintained at each port 12. The maximum number of bins, "P", is chosen based on various factors, such as the amount of available storage space, the frame reception rate, the rate of re-synchronization of the forwarding table contents as mentioned above, etc. Each bin 46 can store an address and a count. A bin 46 is used to store an address that has been determined to be unknown at the port and a count of the number of times frames containing that address have been flooded to the other ports 12 of the network element 10. The bins are allocated and utilized as described below with reference to Figure 5.

Figure 5 shows the process of monitoring flooding and conditionally transferring information among the ports 12 for updating the port forwarding tables. When a unicast frame received at a port 12 is determined to be unknown and is therefore flooded to the other ports 12, it is determined at step 48 whether



ATTORNEY DOCKET NO. FJFNX-183XX
WEINGARTEN, SCHURGIN,
GAGNEBIN & LEBOVICI LLP
TEL. (617) 542-2290
FAX. (617) 451-0313

the DA of the frame has been "binned", i.e., occupies an active bin 46 with a corresponding count. If not, then at step 50 it is determined whether the maximum permissible number, P, of bins 46 are currently active. If not, then at step 52 a new bin 46 is created. The DA of the frame is written to the address portion of the bin 46, and the count is set to 1.

If at step 50 the maximum permissible number of bins 46 are determined to be active, then at step 54 a process for full re-synchronization of the forwarding tables of the ports 12 in the same aggregated port is performed. As a result of such re-synchronization, some or all of the addresses in the active bins 46 become known at the port. The active bins 46 are de-activated or discarded. If the address remains unknown at the port despite the re-synchronization, then it will likely be placed in a new bin 46 the next time a frame containing the address is flooded and the process of Figure 5 is performed.

If at step 48 it is determined that a bin 46 for the unknown address is already active, then at step 56 it is determined whether the count in the bin 46 has reached a predetermined maximum value. This value is a parameter chosen to reflect a desired balance between undesirable frame flooding potentially unnecessary inter-port data transfer. If the maximum value is set too high, then frames containing this address will be flooded numerous times before action is taken to update the forwarding table. If the maximum value is set too low, inter-port data transfers may be requested unnecessarily for addresses that would otherwise be learned in due course via the regular egress learning mechanism.

If at step 56 the maximum count value has not been reached, then at step 58 the count is incremented. This action keeps track of the number of times frames containing this address have been flooded.



ATTORNEY DOCKET NO. FJFNX-183XX
WEINGARTEN, SCHURGIN,
GAGNEBIN & LEBOVICI LLP
TEL. (617) 542-2290
FAX. (617) 451-0313

If at step 58 the maximum count value has been reached, then at step 60 it is determined whether this address is known at any of the other ports 12. If so, information identifying the port 12 associated with this address is transferred from such other port to the port 12 at which the address is unknown, and this port 12 uses the information to create a new entry in its forwarding table. The bin 46 for this address is then discarded.

It is preferable that there be a limit to how long each bin 46 can remain active, to help ensure that the bins 46 are used primarily for active traffic streams rather than inactive or very low-rate streams. If the time limit is exceeded before the count reaches the maximum value that triggers the inter-port information transfer of step 60, the bin is aged out and discarded. Such a time limit is preferably substantially less than the period of full re-synchronization of the forwarding tables, but should be sufficiently high to enable efficient use of the bins 46 for active streams.

It will be apparent to those skilled in the art that modifications to and variations of the disclosed methods and apparatus are possible without departing from the inventive concepts disclosed herein, and therefore the invention should not be viewed as limited except to the full scope and spirit of the appended claims.



ATTORNEY DOCKET NO. FJFNX-183XX
WEINGARTEN, SCHURGIN,
GAGNEBIN & LEBOVICI LLP
TEL. (617) 542-2290
FAX. (617) 451-0313